



基于全基因组信息的“柑”“橘”定义、分类与演化

徐强^{†*}, 黄跃[†], 邓秀新^{*}

华中农业大学, 果蔬园艺作物种质创新与利用全国重点实验室, 武汉430070

[†] 同等贡献

* 联系人, E-mail: xuqiang@mail.hzau.edu.cn; xxdeng@mail.hzau.edu.cn

收稿日期: 2023-11-28; 接受日期: 2024-02-07; 网络版发表日期: 2024-03-11

国家自然科学基金(批准号: 31925034)、国家重点研发计划(批准号: 2022YFF1003100)、中央高校基本科研业务费专项基金(批准号: 2662022PY006)和湖北省博士后创新研究岗位资金(批准号: 307812)资助

摘要 柑橘是全球广泛种植的经济作物之一。柑橘种间甚至属间可以杂交, 且无性系变异普遍, 形成多样性高、遗传关系复杂的物种群。柑橘的分类也成为植物科学难点。本研究收集迄今为止种类最全面的柑橘全基因组数据, 对各种柑橘类型的特点、遗传成分与比例进行精确的分析, 把橘、柚及其杂种分成5个类型, 并进行定义。以“沃柑”和“红美人”两个品种为例解析柑橘杂交历程和遗传组分变化, 并推测橘从驯化之后的传播路线。本文对柑橘物种分类和起源地、“柑”和“橘”的遗传异同、以及橘的传播与演化进行讨论。研究结果基于全基因组信息提出一个比较客观的柑橘分类, 为种质资源的利用提供框架。

关键词 柑橘, 分类, 起源, 基因组

柑橘是全球性大宗水果之一, 在140多个国家和地区有种植, 我国种植面积和产量均居第一位(<https://www.fao.org/faostat/>)。由于柑橘属种类多、芽变频率高、不同种甚至不同属之间存在杂交亲和性, 遗传背景十分复杂, 柑橘家族的起源和分类成为植物学广泛关注的热点问题之一。中国是柑橘起源地之一, 品种繁多, 野生资源极为丰富。枳属、金柑属和柑橘属均原产于我国^[1,2]。我国记录最早的柑橘出现在《禹贡》一书, “厥包橘柚, 锡贡”指的是包裹的橘和柚为大禹王的贡品。澳大利亚本土有澳洲指檬、澳洲沙檬、多蕊橘等野生种^[3], 印度东北部也报道出印度野橘和枸橼等野生种^[4], 这些地方均被不同观点认为是柑橘起源

中心。

柑橘属于芸香科(Rutaceae)、柑橘亚科(Aurantioideae)、柑橘属(*Citrus*)。南宋韩彦直所著《橘录》是世界上最早关于柑橘的专业书籍, 其中记录柑8种, 橘14种, 橙5种。林奈(Linnaeus)于1753年定义柑橘属, 在过去270年里, 研究人员基于形态学、同工酶、化学成分、细胞学和分子标记等分类方法对柑橘属物种进行分类, 得到多种柑橘分类系统, 有些分类系统甚至截然不同。目前有2套分类系统比较为国际所接受, 分别是美国分类学家施文格提出的包含16个种的大种分类系统^[3]和日本分类学家田中长三郎提出的包含162个种的小种分类系统^[5]。施文格分类系统包括柑橘亚属(10

引用格式: 徐强, 黄跃, 邓秀新. 基于全基因组信息的“柑”“橘”定义、分类与演化. 中国科学: 生命科学, 2024, 54: 525-536
Xu Q, Huang Y, Deng X X. The definition, classification and evolution of citrus based on whole genome sequence information (in Chinese). Sci Sin Vitae, 2024, 54: 525-536, doi: 10.1360/SSV-2023-0229

个种)和大翼橙亚属(6个种), 是目前接受度最广的系统. 另外一个有影响力的研究是Scora提出柑橘属主要由三个基本种构成, 分别是橘(*Citrus reticulata*), 柚(*C. maxima*)和枸橼(*C. medica*); 而葡萄柚、柠檬和甜橙等都是由以上3个基本种杂交而来, 这个结论获得广泛的支持^[6-15]. 后来Mabberley^[16]提出食用柑橘可分成基本种和杂交品种群; 并提出枳属、金柑属、澳指檬属、澳沙檬属和多蕊橘属应该下降到种的地位并合并到柑橘属, 这样柑橘属包括11个基本种和野生种, 以及14个代表性的杂交种^[17]. 我国柑橘分类学家曾勉^[18]结合柑橘属地理分布和形态结构特点, 将柑橘分为大翼橙属、枸橼属、柚属、柑橘属和橙属五个属, 其中柑橘属共包含30个种. 现有分类系统大多数基于表型^[19]、生化^[20]、同工酶^[12,13]、染色体核型^[21]、DNA条形码^[22]、分子标记^[8]、基因序列^[23]等信息, 这些分类系统为揭示芸香科和柑橘属植物的遗传关系提供重要线索. 全基因组数据的出现为全面系统地研究植物分类提供可靠的信息^[24-28].

橘是柑橘中最重要的基本种, 大多数柑橘杂交品种中都有来自橘的遗传成分. 中国南岭山脉存在大量野生橘的自然群体, 莽山野橘表现出非常纯的遗传背景, 随后野橘在南岭山脉的南北两侧分别发生两次独立驯化事件, 形成现代栽培柑橘^[26,29]. 此外, 在琉球群岛上也发现橘的亚种, 这些亚种很可能是在琉球群岛尚未被海水淹没前从中国南部传播过去^[30]. 在传播过程中, 橘基因组中发生柚的基因渐渗, 形成杂柑. “柑”的概念由来已久, 在中国古代, “橘”指小果宽皮柑橘, 而“柑”代指比橘大, 但比柚小的柑橘, 且一般认为“柑”的风味好于“橘”. 在现代, 由于出现更多种类的柑橘类水果, “橘”和“柑”的概念开始混淆. 广义上, “橘”指代砂糖橘、南丰蜜橘等果实较小、易剥皮的品种; “柑”通常指温州蜜柑、贡柑等品种, 这些品种一般为橘和甜橙的杂种后代, 因此它们的基因组中有来自柚的基因渐渗, 其果实大小比“橘”大, 且剥皮难度大于橘. 随着品种数量增多, 呈现出“橘橙”、“橘柚”等新的分类名称, 因此传统的分类方法很难对所有的品种进行准确的归类. 通过基因组测序分析可以对传统的“柑”和“橘”进行精准分类, 王小柯等^[29]利用240份橘简化基因组测序信息(genotyping-by-sequencing, GBS)将橘分为11个类群, 且不同地理来源的宽皮柑橘在遗传水平上存在显著差异. Wu等人^[31]通过解析橘基因组

中来自柚的基因渐渗比例将橘分为三种类型, 第一种类型的橘为纯种橘, 出现的时间最早, 第二种类型的橘中存在一定比例的柚基因渐渗, 出现的时间晚于第一种类型, 而第三种类型的橘中来自柚的基因渐渗的比例最高, 出现的时间最晚. 本研究通过对大量橘柚杂种进行测序, 对橘柚杂种进行更精细和较完善的分类.

1 方法

1.1 序列比对

对分析样品进行第二代测序, 测序平台为Illumina. 使用fastp v0.23.2^[32]对原始测序数据进行质量过滤, 随后用Burrows-Wheeler-Alignment Tool(BWA v0.7.17)^[33]将过滤后的序列比对到甜橙参考基因组, 得到sam格式的比对文件. 用Samtools v1.14^[34]将比对的结果转化为bam格式并进行排序, 并使用Picard(<https://github.com/broadinstitute/picard>, v2.72.2)和GATK v3.8.1^[35]去除测序中产生的PCR冗余, 得到最终的比对结果, 并鉴定SNP.

1.2 杂合度分析

设置100 kb大小的窗口, 统计窗口中杂合SNP的数量, 将其与窗口大小的比例定为该窗口的杂合度. 扫描全基因组序列并依次计算所有窗口的杂合度, 并编写R脚本将全基因组所有窗口的杂合度以小提琴图进行展示.

1.3 系统进化分析

分别使用串联模型和溯祖模型进行系统进化分析. 使用串联模型时, 提取位于四兼并碱基位点的SNP, 将提取的SNP信息转化为Phy格式, 使用modeltest-ng v0.1.7^[36]选择最合适的进化拟合模型, 随后使用raxml-ng v1.1^[37]构建系统进化树, 选择的模型为“TVM+I+G4”, 将酒饼筋作为外类群, Bootstrap值设为1000. 根据在临沧发现的柑橘属化石, 将柑橘属出现的时间设为八百万年前, 并以此时间为基准, 推测柑橘属其他物种的分化时间, 使用软件为R8s v1.8.1^[38].

使用溯祖模型进行分析时, 先用raxml-ng v1.1分别对来自不同基因的SNP构建基因树, 随后使用Astral v5.7.8^[39]对这些基因树进行分析, 构建柑橘属的物种

树, 并计算物种树的quartet score. 构建叶绿体进化树时, 将测序的序列比对到甜橙叶绿体基因组, 提取所有的SNP位点, 并利用raxml-ng v1.1构建进化树.

1.4 基因组成分分析

基因组成分鉴定的方法参考Wang等人^[26]的研究, 并进行一定的修改, 具体流程如下: 将基本种的基因组进行比较, 鉴定每个种特异的纯合SNP, 将这些SNP称为诊断型SNP, 并记录诊断型SNP的位点和基因型. 在鉴定一个未知物种时, 分析该物种在诊断型SNP所在位置的基因型, 若基因型和基本种一致, 则认为未知物种该位点可能来源于对应的基本种. 以100 kb大小的窗口扫描全基因组, 若一个窗口中, 大于50%的诊断型SNP来自一个基本种, 则认为该窗口的基因组成分为该基本种, 否则该窗口的成分为未知. 当检测窗口的基因组来源为两个基本种时, 该基因组的成分为这两个基本种的杂种, 若检测窗口的基因组来源大于两个基本种时, 则认为基因组成分未知.

1.5 基因流分析

分别使用treemix v1.13^[40]和ABBA-BABA v1.0^[41]分析来评估橘、柚和它们杂交后代之间基因流. 在两个分析中, 均使用莽山野柑作为外类群. 在进行ABBA-BABA v1.0分析时, 用橘作为P1, 柚作为P3, 分别用柑、橙和葡萄柚作为P2.

2 结果

2.1 柑橘属的物种特点

本研究共获得147份柑橘全基因组测序数据, 其中包括34份橘、38份柑、3份葡萄柚、12份橙、5份柚、15份枸橼和柠檬、4份金柑及其杂交后代、12份枳及其杂交后代、14份宜昌橙及其杂交后代, 此外, 还收集10份其他柑橘基本种和酒饼筋的测序数据(网络版附表1). 21份数据为本研究新测的数据. 本文前期基于314份柑橘亚科的基因组研究, 对柑橘属的物种进行界定并提出15个种的假说^[42]. 本研究分别使用串联模型和溯祖模型对柑橘属的物种进行系统进化分析, 两种模型分析得到的系统进化关系一致(图1A和网络版附1). 进化树结果支持原先提出的假说并进一步丰富物种信息. 柑橘属可以分为三个物种群. 物种

群I包括橘-临武野橘, 该物种群主要分布在南岭山脉北麓以及附近区域, 与另两个物种群相比, 该物种群和早期分化的柑橘物种(如莽山野柑和宜昌橙)遗传关系最近, 且分化时间早于另两个物种群(图1A). 物种群II是马蜂柑-柚-枸橼物种群, 该物种群主要分布在喜马拉雅山东南麓, 其中包括马蜂柑、柚、红河大翼橙、枸橼和印度野橘, 其中柚和红河大翼橙以及枸橼和印度野橘分别具有较近的亲缘关系. 物种群III是金柑-澳大利亚物种群, 包括金柑、澳洲沙檬、澳洲指檬、澳洲圆来檬、多蕊橘等物种.

柑橘属物种的性状多样性丰富, 叶、花、果呈现出不同的特征(图1B). 枳(*Citrus trifoliata*)果实表面附有绒毛, 味苦, 常作为中药使用, 其种子多胚, 叶片为三出复叶. 枳原产于中国, 具有良好的抗逆性, 是常用的柑橘砧木. 莽山野柑(*C. mangshanensis*)是原产于中国湖南莽山地区的野生柑橘, 近期在广西也发现和莽山野柑非常类似的野生资源(网络版附2). 莽山野柑属于柑橘属内一个独立的种, 而非橘柚杂交形成的“柑”, 其果实表皮褶皱, 成熟后具有特殊的香气. 宜昌橙(*C. ichangensis*)首次在湖北宜昌被发现, 在湖北、湖南、重庆、广西、云南等地均有分布, 宜昌橙果实较小、果皮较厚, 不可食用, 其树枝多刺, 叶片的翼叶较大, 花通常为白色或粉色. 在物种群I, 临武野橘(*C. linwuensis*)是近期在湖南莽山附近发现的野生柑橘, 基因组分析结果显示, 它属于柑橘属的一个新种^[42], 它的嫩叶呈紫色, 枝条偏软, 果实香味明显, 种子多. 在物种群II, 马蜂柑(*C. hystrix*)是分布在云南、马来西亚等地区野生资源, 其果实表面不平整, 翼叶较大接近叶身大小, 果实和叶片均存在特殊的香气. 红河大翼橙(*C. hongheensis*)是最初发现于云南红河县的野生柑橘, 果实和柚很接近, 为圆形或扁圆形, 果皮较厚, 果味酸、微苦. 叶片狭长, 翼叶非常明显, 其长度通常为叶身长度的2~3倍. 印度野橘(*C. indica*)是原产于印度的野生资源, 它是柑橘属中独立的种, 并不属于传统认为的橘, 它的果实较小, 果皮光滑, 种子较大. 山金柑(*C. hindsii*)的叶、花、果器官均较小, 果皮为橙色, 可作为观赏植物; 其开花结果快, 正在被开发为研究柑橘遗传和基因功能的模式植物. 在物种群III, 澳洲沙檬、澳洲指檬和多蕊橘起源于澳大利亚及附近岛屿, 其中, 澳洲沙檬(*C. glauca*)生长在沙漠地带, 具有较强的耐旱性, 其果实较小, 叶片呈长条状. 多蕊橘(*C. poly-*

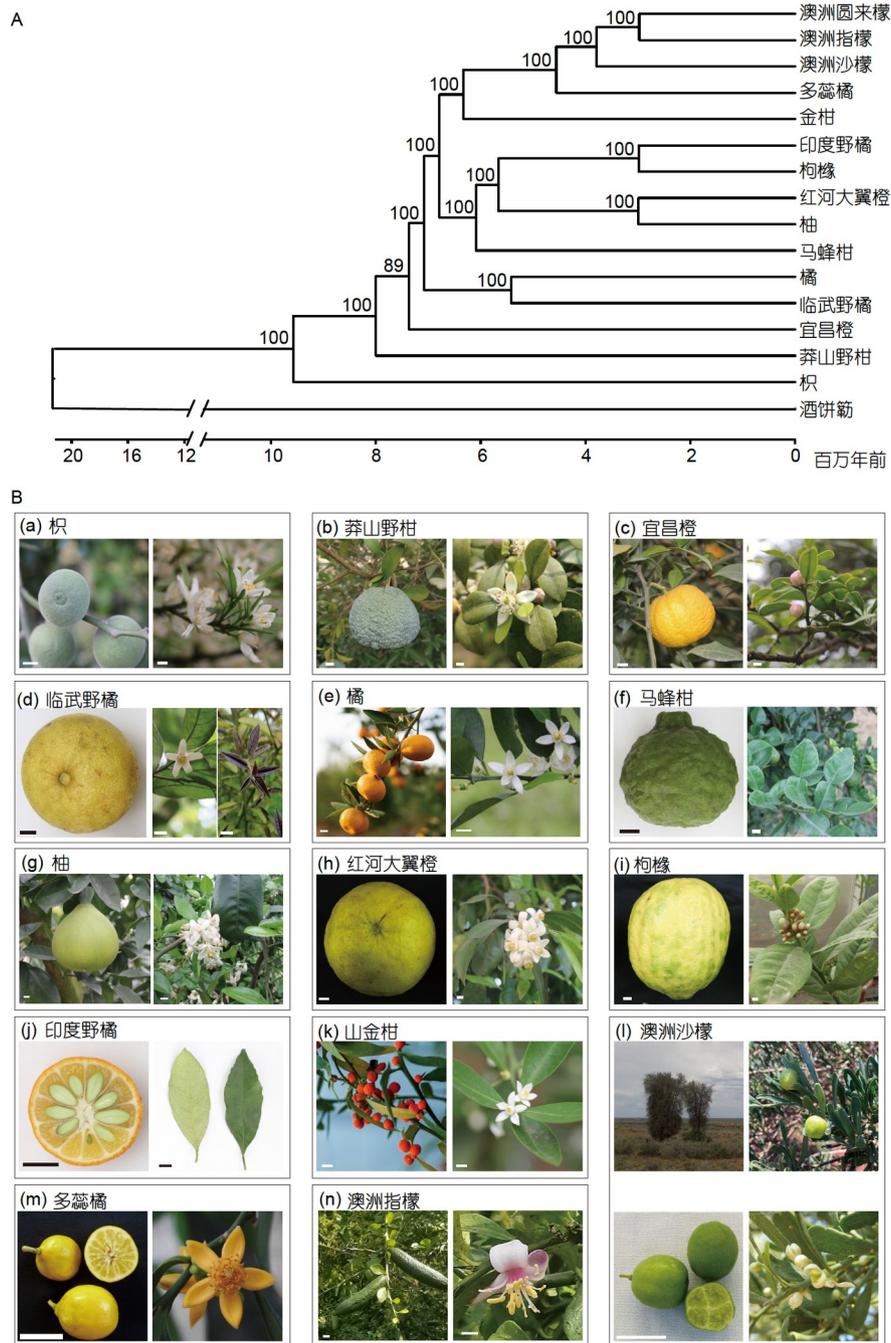


图 1 柑橘属物种的系统进化树和叶、花、果特征。A: 柑橘属物种的系统进化树。酒饼筋作为外类群。下方的标尺表示物种的分化时间, 单位为百万年前; B: 柑橘属物种的叶、花、果特征。(a)枳的幼果和花; (b)莽山野柑未成熟果实和花; (c)宜昌橙成熟果实和花; (d)临武野橘成熟果实、花和叶; (e)橘(南丰密橘)的果实和花; (f)马蜂柑果实和叶片; (g)柚果实和花; (h)红河大翼橙果实和花; (i)枸橼果实和花; (j)印度野橘果实和叶片; (k)山金柑果实和花; (l)澳洲沙檬生长环境、叶、果实和花; (m)多蕊橘果实和花; (n)澳洲指檬果实和花。标尺为1 cm。

Figure 1 The phylogenetic tree and characteristics of leaves, flowers and fruits of *Citrus* species. A: The phylogenetic tree of *Citrus*. *Atalantia buxifoliata* was used as outgroup. The scale at the bottom indicates the divergence time in million years ago; B: characteristics of leaves, flowers and fruits of *Citrus*. Young fruits and flowers of *C. trifoliata* (a) and *C. mangshanensis* (b); (c) mature fruits and flowers of *C. ichangensis*; (d) mature fruits, flowers and leaves of *C. linwuensis*; fruits and flowers of *C. reticulata* (nanfengmiju) (e); fruits and leaves of *C. hystrix* (f); *C. maxima* (g); *C. hongheensis* (h); *C. medica* (i); fruits and leaves of *C. indica* (j); *C. hindsii* (k); (l) the ecological environment, leaves, fruits and flowers of *C. glauca*; *C. polyandra* (m) and *C. australasica* (n). Scale bars, 1cm

andra)原产于新几内亚,它的果实呈橙黄色,囊瓣数量较多,其花瓣颜色特别为亮橙色(图1B和M).澳洲指檬(*C. australasica*)果实形似手指,并包含颜色鲜艳的汁胞,其叶片呈较小的椭圆形,花瓣为淡紫色或白色.

对现有杂交柑橘品种有重要贡献被广泛认为是三个基本种.橘(*C. reticulata*)是最常见的三个基本种之一(图1B和E),其中包含大量的商业品种,橘的翼叶较小,果实大小适中,易剥皮,因此又被称为“宽皮柑橘”.柚(*C. maxima*)也是常见的三个基本种之一(图1B和G),其果实大,皮厚,果肉呈红色或白色,相较于其他柑橘,柚的叶片较大,翼叶明显,花呈簇状分布.枸橼(*C. medica*)是柑橘基本种的最后一种(图1B和I),它的果实较大,白皮层较厚,部分品种的果实没有汁胞,枸橼的叶片呈椭圆形,无翼叶,花蕾为紫红色.

2.2 “柑”“橘”的基因组成份与分类

柑橘属间存在广泛的种间杂交,导致柑橘属遗传背景高度复杂.在15个物种中,橘(*C. reticulata*)和柚(*C. maxima*)之间的杂种是种类最多的柑橘杂种,但由于习惯,一般都统称为柑橘(俗称柑橘,还包含柚子等).本研究分析柑橘属物种的基因组杂合度,结果显示甜橙、酸橙、葡萄柚等典型杂种的杂合度明显高于基本种,而“柑”的杂合度大多位于基本种和典型的杂种之间,这表明柑橘的杂种基因组十分复杂(网络版附图3).本研究进一步使用treemix v1.1.3和ABBA-BABA v1.0来评估基本种和杂种之间的基因流.两种方法均检测到从柚到柑、橙和葡萄柚的基因流,而且到葡萄柚和橙的基因流比例要高于到柑的基因流比例(网络版附图4和5).

育种人员通常利用表型(如是否易剥皮)或根据亲本类型来对柑橘杂种进行分类,但许多杂种亲本未知,而仅靠表型难以对杂种进行精准分类,这导致柑橘杂种的分类十分混乱.为克服这个难题,本文从全基因组水平鉴定每个基本种特异的纯合SNP,将这些SNP作为鉴定遗传背景的“诊断型SNP”(详见方法).利用这些诊断型SNP来鉴定柑橘基本种和杂交种的遗传背景.本研究共鉴定到5279839个诊断型SNP,具体每个物种鉴定到的诊断型SNP数据见网络版附表2.

根据柑橘基因组中橘、柚所占的比例,本文将橘、柚及其杂种分成5大类型(图2和网络版附表3).橘柚杂种是栽培柑橘种最常见的类型,其基因组中存在

橘纯合(橘/橘),柚纯合(柚/柚)以及橘柚杂合(橘/柚)三种不同类型的基因组成分(图2).当基因组橘/橘比例达到80%以上定义为橘,基因组的柚/柚比例达到80%定义为柚.当橘/橘的比例低于80%,且渗入柚/柚或橘/柚的比例均低于80%时,定义为柑,代表品种包括克里曼丁和温州蜜柑等.当橘/柚的比例达到80%以上时,定义为橙(*C. sinensis*),其代表品种包括血橙、脐橙、酸橙等.而当柚的比例位于20%至80%之间且不存在橘/橘时,定义为葡萄柚(*C. paradisi*).在不同的橘柚杂种中鉴定到的来自柚的比例也符合基因流分析的结果.

除橘柚杂种外,枸橼等其他物种也可以和橘、柚形成杂种.柠檬由枸橼和酸橙杂交形成,其基因组包含枸橼和橘的杂合成分(枸橼/橘)以及枸橼和柚的杂合成分(枸橼/柚).此外,小香橼和红来檬的基因组绝大部分为枸橼/橘杂合(网络版附表4).四季橘的基因组成分包含金柑/橘和金柑/金柑(网络版附表5).枳能与橘、柚、甜橙杂交,分别形成枳橘、枳柚和枳橙.枳橘的基因组绝大部分为枳/橘杂合,枳柚的基因组绝大部分为枳/柚杂合,而枳橙的基因组包含枳/橘和枳/柚两种成分(网络版附表6).香橙(又叫日本香柚)基因组绝大部分为宜昌橙/橘杂合(网络版附表7).

2.3 橘品种群在我国的传播

橘起源于我国南岭山脉一带,随后分别往南北两个方向传播,并形成两个种群^[26].往南传播的橘主要分布在广东、广西等地,代表性品种包括砂糖橘、茶枝柑、马水橘等.往北传播的代表性品种分布在湖南、四川、江西等地,包括红橘、丹橘和南丰蜜橘等.基因组分析结果显示橘在传播过程中渗入橘/柚的基因组成分,这些基因的渗入可能导致野橘到栽培橘的驯化.本研究进一步分析由莽山往南及往北传播的两个类群的橘,系统进化分析结果显示这两个类群的橘被聚为两个分支,基因组成分表明这两类橘中橘/柚渗入的基因组区域存在一定差别,往北传播的橘大多数在2号染色体的前端、8号染色体的末端及9号染色体的大部分区域发生渐渗,而往南传播的橘则大多在6号染色体的前端和8号染色体的大部分区域发生渐渗(图3A).

往北传播的橘随后往东传播,并通过杂交和芽变选种形成一批具有地方特色的品种.例如乳橘和南丰



图 2 基于基因组成分的橘、柚及其杂种分类. 每行代表一个种类, 第一列和第二列为种类名称和代表品种果实图片, 第三列柱形图为代表品种的基因组成分, 第四列为根据基因组成分进行物种分类的判断标准, 第五列为该类的代表性品种. 标尺为 1 cm.

Figure 2 The classification of mandarin, pummelo and their hybrids based on their genetic compositions. Each line defines the classification of a category. The first and second columns show the name of the category and a picture of the fruit of the representative species. The third column shows the genomic components of representative species. The fourth column shows the criteria for defining the category based on genomic composition. The fifth column shows the names of the representative cultivars of the category. Scale bars, 1cm

蜜橘聚在一起, 且基因组成分一致, 表明乳橘可能由江西传往浙江. 浙江本土特色品种满头红为朱红橘的芽变, 可能最初由长江传播并经选育而来. 此外, 部分国外品种和中国的品种具有一致的基因组成分, 这些国外品种可能最初源于中国, 到国外后经选育形成新品种. 如地中海附近的柳叶橘可能源于茶枝柑, 而温州蜜柑也最早来源于浙江.

在橘传播过程中, 来自于柚的基因渐渗和果实品质紧密相关. 果实形态和果实风味是决定柑橘品质的重要因素, 本研究分析不同基因组成分的比例和性状之间的关系. 结果显示, 基因组中柚成分的比例和果皮厚度及果实重量呈显著正相关, 相关系数分别为 $0.818 (P < 0.0001)$ 和 $0.814 (P < 0.0001)$ (图3B). 本文进一步分析橘柚杂种中影响风味的代谢物质的含量, 包括陈皮素、柚皮素、柠檬苦素和诺米林. 结果显示, 橘柚杂种基因组中柚成分比例和陈皮素的含量呈显著负相关, 相关系数为 $-0.774 (P < 0.0001)$, 而和柚皮素、柠檬苦素和诺米林的含量呈显著正相关, 相关系数分别为 $0.721 (P < 0.0001)$ 、 $0.539 (P = 0.0014)$ 和 $0.661 (P < 0.0001)$ (网络版附图6).

2.4 杂交柑橘的改良历程

杂交是现代柑橘育种的一个重要手段, 部分柑橘

品种也常作为重要的杂交亲本, 如椪柑、红橘、甜橙等. 红美人和沃柑是近期在市场上备受欢迎的两个杂种, 本研究以这两个品种为例构建主要品种的杂交路线, 并解析重要杂种后代的基因组成分(图4). 在这个杂交路线中, 酸橙和甜橙分别是第一代和第二代杂种. 甜橙处于关键的育种位置, 它作为所有第三代杂交子代的亲本, 包括本地广橘、庙宇柑、克里曼丁和葡萄柚. 沃柑和红美人分别是第四代杂种和第八代杂种(图4). 值得指出的是第1~3代基本上是天然杂交, 第4代是人工杂交和自然杂交的过渡期, 第4代之后是人工杂交. 柑橘人工杂交仅有100年左右的历史.

利用叶绿体基因组序列, 推测杂交历程中的母本来源(网络版附图7). 结果显示, 克里曼丁、庙宇柑、天草、红美人、温州蜜柑、清见、南香、沃柑聚在一起, 且它们的叶绿体最初均来源于橘; 柠檬、酸橙、甜橙、本地广橘、葡萄柚、明尼奥拉、佩奇橘聚在一起, 它们的叶绿体最初均来源于柚. 此外, 解析这些主要杂交柑橘的基因组背景, 除葡萄柚外, 第三代杂交子代和第四代杂交子代均使用橘作为亲本, 因此他们基因组成分均为橘/橘和橘/柚, 而在第五代及之后的杂交子代中, 基因组成分中出现柚/柚.

杂交造成的基因组成分的改变会影响子代的性状, 对基因组成分的解析有利于预测子代的性状, 从而

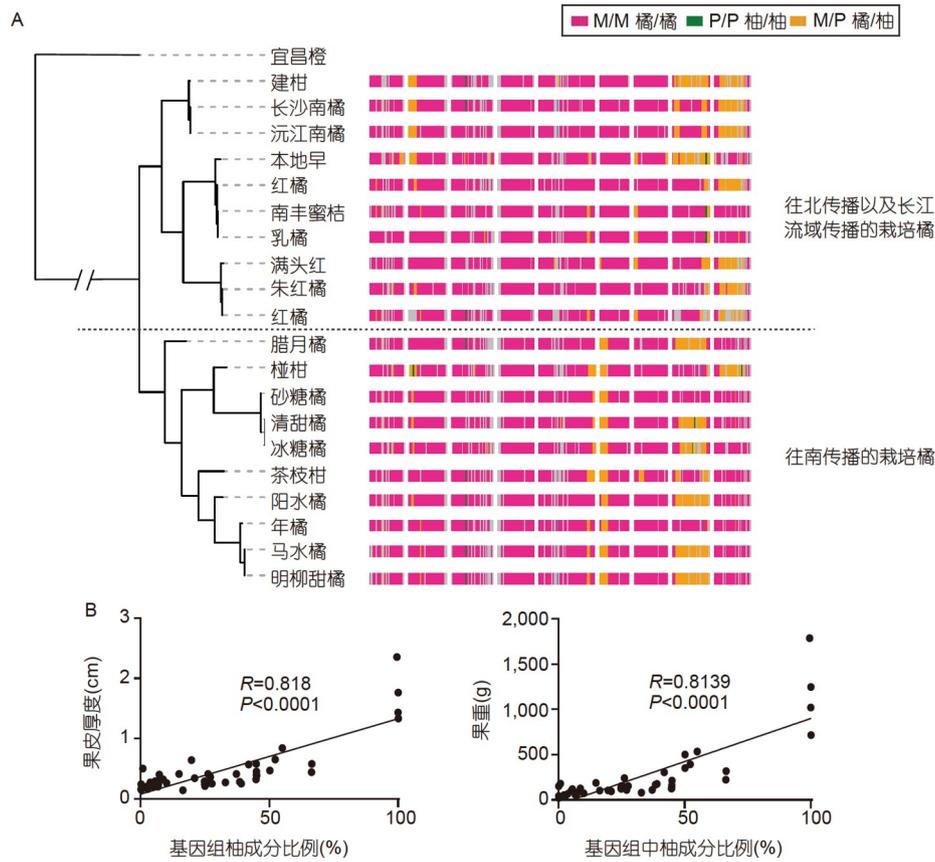


图 3 橘及其杂种群基因组特征. A: 往北和往南传播的两类橘的系统进化树和基因组成分; B: 橘柚杂种基因组中柚成分的比例和果皮厚度及果重之间的相关性

Figure 3 The genomic characterization of mandarin and its hybrids. A: Phylogenetic tree and genomic compositions of two groups of mandarin that spread northward and southward; B: correlations between the genomic proportion of pummelo and peel thickness and fruit weight in the hybrid citrus

进行辅助育种. 多胚性是柑橘生殖中的一个重要性状, 多胚柑橘有利于维持优良性状. 前期研究显示 *RWP* 基因启动子区的一个 MITE 转座子插入与柑橘的多胚性紧密连锁^[28]. 该 MITE 插入广泛存在于橘, 而不存在于柚. 基于基因组成分析的结果, 在本研究的 53 个橘柚杂种后代中, 本文在 12 个杂交后代中没有鉴定到该 MITE 的插入, 表明这 12 个杂交后代均为单胚, 可作为杂交育种的候选亲本(网络版附表 8).

3 讨论

3.1 柑橘属的物种分类与起源地

柑橘分类一直众说纷纭, 从最初林奈提出的 2 个种到 Tanaka 系统的 162 个种^[5], Swingle 分类系统提出 16 个种^[3], Hodgson 系统则包含 39 个种^[43]; Mabberley 提出 25

个种^[17]; 我国曾勉^[18] 提出的分类系统总共承认 30 个种. 这些分类系统都是基于表型或者生化数据, 存在各自的优缺点, 且都有使用者和追随者. 近期基于柑橘亚科基因组研究提出柑橘属包括 15 个种, 并分为一个早期分化的群以及三个包含栽培品种的物种群^[42]. 本研究整合更多的数据, 系统进化分析结果也支持该分类, 并对杂交柑橘进行更精细的剖析. 与此前广泛认同的柑橘属主要由三个基本种(橘、柚和枸橼)相比较, 橘和物种群 I 吻合, 柚和枸橼均属于物种群 II, 而澳大利亚物种并没有被充分体现. 作者认为, 柑橘属应该有 4 个基本种, 分别是橘、柚、枸橼和澳洲指檬. 目前商业品种主要是橘和柚杂交种, 本研究进一步明确分为 5 个类型, 即橘、柑、橙、葡萄柚和柚. 此外部分商业品种是橘、柚和枸橼的杂交种, 如柠檬.

早期分化的柑橘野生物种以及物种群 I 主要分布

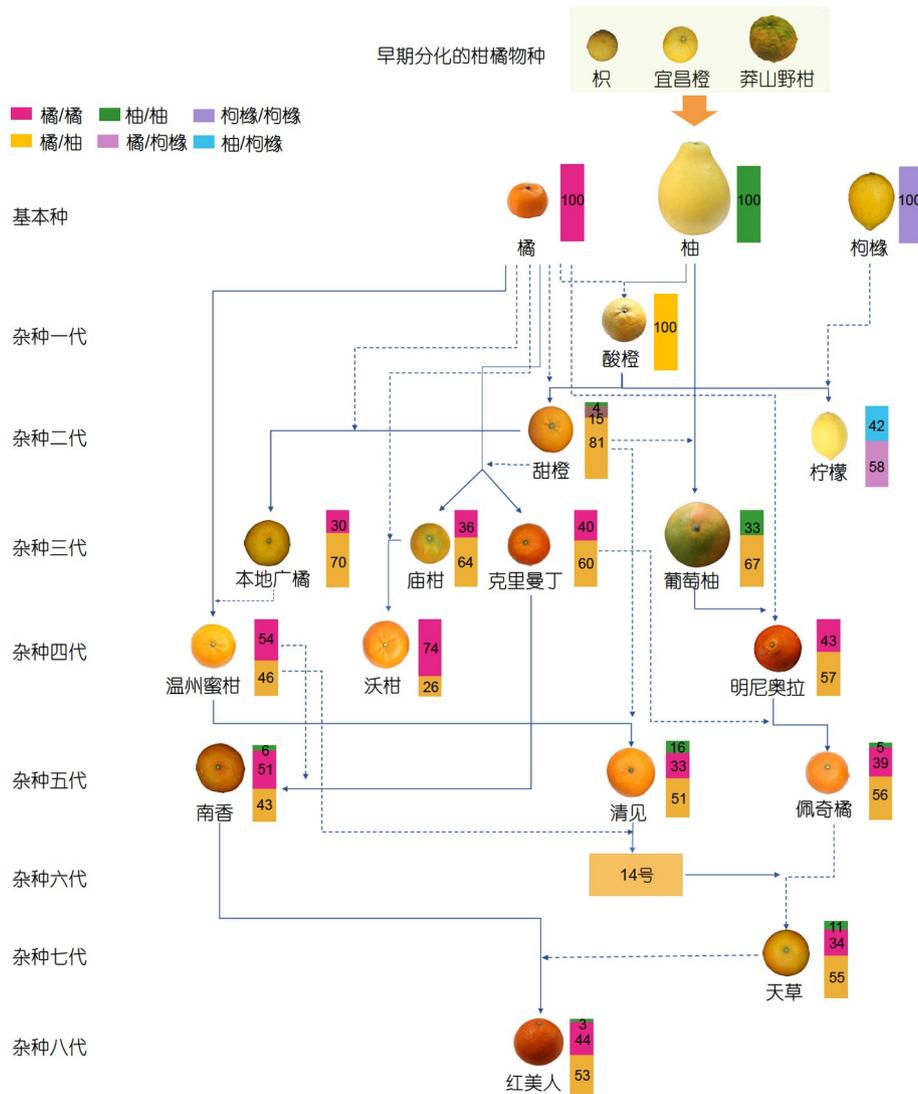


图 4 以沃柑和红美人为代表解析主要杂交柑橘品种的育种路线。果实旁边的竖条表示核基因组的遗传构成，内部的数值表示不同遗传成分所占的比例。左上角不同颜色小长方形表示不同的遗传成分。浅黄色长方形表明未进行测序的杂交子代14号

Figure 4 Breeding routes for the main hybrid cultivars of citrus represented by the Orah and the Beni Madonna. The bars near the fruits indicate the genetic composition of the nuclear genome, and the values inside the bars indicate the percentage of the corresponding genetic components. Different colored rectangles in the upper left corners indicate different genomic components. The light yellow rectangle with No.14 inside indicates that the genome information of is not available for analysis

在华中地区，暗示着华中地区可能是柑橘属的初生起源中心^[42]。早期分化的物种包括枳、莽山野柑和宜昌橙。在我国，湖北和河南枳的遗传多样性较高^[44]；近期，在华中地区也发现冬季不落叶的枳及其杂种后代^[42,45]。叶绿体序列及基因组序列分析表明，枳和柑橘属形成一个单系类群^[23,46]，但枳和柑橘属核DNA差别比较明显，枳可能是近缘种向柑橘属的过渡种，具有较独特的进化地位。莽山野柑和宜昌橙均为常绿树

种，与柑橘均有杂交亲和性，且杂种后代可育。莽山野柑特异分布在南岭山脉北麓莽山一带，宜昌橙在我国南方地区有广泛分布但以华中地区的多样性高^[47]，均支持早期柑橘的起源中心在华中地区。另外以橘为代表的物种群 I 分化最早，主要起源于该地区，且存在各种野生橘的天然群体。

一直以来，柑橘属被认为起源于喜马拉雅山脉东南麓，包括中国云南、缅甸北部以及印度东北部区域。

本研究认为, 以柚和枸橼为代表的物种群II主要起源于该地区, 该地区也存在各种野生资源, 可以被认为是柑橘属的次生起源中心之一. 澳大利亚的柑橘资源一直也存在较多争论, 本研究结果也表明, 澳大利亚也是柑橘的次生起源中心之一.

综合本研究以及近期关于柑橘基因组和起源的研究^[42], 柑橘丰富的种质资源主要形成一个初生起源中心(中国中华华南地区)和两个次生起源中心(喜马拉雅山脉东麓和澳大利亚地区).

3.2 “柑”和“橘”的定义和分类

宽皮柑橘的变异类型复杂, 也是争论最为激烈的分类. 钟广炎和叶荫民^[14]表明宽皮柑橘存在两大类型, 一是“柑”, 包括温州蜜柑、王柑、蕉柑等类型, 二是“橘”, 如砂糖橘、南丰蜜橘、椪柑等容易剥皮的类型. “柑”的形态介于橘和甜橙之间, 一些类型如蕉柑、温州蜜柑是橘橙杂种, 本研究采用“二八分”原则进一步将柑橘的分类精确定义, 认为基因组中橘/橘的比例大于80%的品种为橘, 橘/橘和橘/柚的比例均在0%至80%之间, 且柚/柚的比例低于80%的品种为柑. “橘”被认为柑橘属的基本种之一, 古代认为“橘小柚大”. 同工酶分析表明小果宽皮橘类, 特别是高酸野生类型宽皮橘是原始类型, 和栽培类型遗传关系明显分开^[13], 道县野橘和土橘首先和栽培类型分开, 表现为明显的两大类^[15]. 野橘的遗传背景较纯, 而栽培橘有种间或者古老基因组的渐渗, 导致其复杂性^[48]. 南岭北麓一带以莽山为代表是野橘重要起源地, 莽山野橘(*C. reticulata*)、元橘可能是宽皮橘类的始祖^[27]. 通过分子标记分析表明, 莽山野橘的一种类型与道县野橘和崇义野橘较近, 是栽培柑橘的原始类型^[49]. 通过高通量测序技术揭示240份宽皮柑橘可以划分为11个群, 与地理来源和特定形态紧密相关, 南岭山脉南北的品种群存在不同演化路径^[29]. 通过全基因组分析进一步聚焦, 揭示莽山南北发生两次独立的野橘驯化事件, 产生遗传背景不同的栽培橘, 传播到各地后经杂交形成丰富的

“柑”^[26]. 值得一提的是, 莽山野柑(*C. mangshanensis*)是柑橘属中独立于橘和柚之外的一个野生种, 而非橘、柚杂交生成的“柑”.

3.3 橘的起源传播

在湖南、广西和广东地区, 也是橘发生驯化的地方, 依然保留着莽山野橘、道县野橘、崇义野橘、贺州野橘等各种野生群体, 以及丰富的驯化后的栽培橘资源, 如砂糖橘、马水橘、春甜橘、红橘、茶枝柑等. 在道县还有一种半驯化的品种——滑皮橘, 该品种酸较高, 有部分野生橘的特征, 但整体果皮已经比较光滑, 可以食用. 部分栽培橘(如红橘、朱红橘、南丰蜜橘等)随后往东传到浙江、江苏等地, 并选育出新的品种, 如乳橘、满头红等. 一些国外的品种, 如日本的温州蜜柑、欧洲的柳叶橘等, 最初也源自中国.

基因渐渗是物种引入新的性状的重要驱动力, 在葡萄^[50]、苹果^[51]、西瓜^[52]等果树育种过程中, 基因的渐渗为栽培品种引入新的性状, 促进新品种的形成. 在橘的传播和驯化过程中发生柚的基因渐渗, 从而形成新的品种, 包括栽培橘和各种不同类型的柑. 柚在果皮厚度、果实大小、代谢物等生理指标上和橘存在差异. 柚的基因渐渗, 改变这些性状, 且渗透的比例与果皮厚度、果实大小有显著正相关, 进而形成多态性的柑橘种群. 例如橘果实较小、重量较轻、易剥皮, 而柚果实较大、重量较重、不易剥皮, 橘柚杂交种的性状则位于两者之间, 且随着渐渗比例的增加, 性状和柚更加接近. 此外, 这些丰富的渐渗系, 也可用于定位控制性状的位点, 为后续柑橘的遗传改良奠定基础. 我国是栽培柑橘的起源地, 这也成就很多天然杂交在中国发生, 回溯历史, 酸橙在汉朝才有记载, 到宋朝有甜橙的记载, 葡萄柚最早被描述是清朝. 可以看出, 在没有人工干预下, 前面一代经历的时间多的上千年, 少的也有几百年. 第4代杂交以后进程明显加快, 到第8代共100年左右历史, 人类的参与明显加快柑橘遗传改良的步伐和柑橘产业在全球的快速发展.

致谢 感谢广西特色作物研究院邓崇岭研究员提供样品分析, 澳大利亚昆士兰农业和渔业部门Malcolm W Smith提供澳洲柑橘资源的图片, 海南大学热带农林学院王华锋教授提供澳指椪图片, 西南大学柑桔研究所江东老师提供明尼奥拉果实图片. 感谢华中农业大学园艺林学学院王伦博士提供诊断型SNP分析经验.

参考文献

- 1 Zhang D, Mabberley D J. *Citrus*. Flora of China, 2008, 11: 90–96
- 2 Zhang W. Thirty years achievements in *Citrus* varietal improvement in China. In: K. Matsumoto ed. Proceedings of the International Society of Citriculture. Tokyo: International Society of Citriculture, 1982. 1982–1983
- 3 Swingle W, Reece P. The botany of *Citrus* and its wild relatives. In: Reuther, W. et al. eds. The *Citrus* industry, History, World Distribution, Botany, and Varieties. California : University of California. 1967, 190–430
- 4 Kumar S, Nair K N, Jena S N. Molecular differentiation in Indian *Citrus* L. (Rutaceae) inferred from nrDNA ITS sequence analysis. *Genet Resour Crop Evol*, 2013, 60: 59–75
- 5 Tanaka T. Fundamental discussion of *Citrus* classification. *Stud Citrol*, 1977, 14: 1–5
- 6 Barrett H C, Rhodes A M. A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. *Systatic Bot*, 1976, 1: 105
- 7 Handa T, Ishizawa Y, Oogaki C. Phylogenetic study of Fraction I protein in the genus *Citrus* and its close related genera.. *Jpn J Genet*, 1986, 61: 15–24
- 8 Moore G A. Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends Genet*, 2001, 17: 536–540
- 9 Ramadugu C, Pfeil B E, Keremane M L, et al. A six nuclear gene phylogeny of *Citrus* (Rutaceae) taking into account hybridization and lineage sorting. *PLoS ONE*, 2013, 8: e68410
- 10 Scora R. Biochemistry, taxonomy and evolution of modern cultivated *Citrus*. In R. Goren and K. Mendel eds. Citriculture: proceedings of the Sixth International *Citrus* Congress. Tel Aviv: Balaban Publisher, 1988: 277–289
- 11 Torres A M, Soost R K, Diedenhofen U. Leaf isozymes as genetic markers in *Citrus*. *Am J Bot*, 1978, 65: 869–881
- 12 Chen L G, Mitsuo O, Tetsushi H. Studies on taxonomy of *Citrus* by GOT isozyme (in Chinese). *Acta Hortie Sinica*, 1991, 18: 27–32. [陈力耕, 大村三男, 日高哲志. 应用 GOT 同工酶进行柑桔分类的研究. 园艺学报, 1991, 18: 27–32]
- 13 Feng L, Song Y, Guan Q, et al. Long-term ethanol exposure inhibits glucose transporter 4 expression via an AMPK-dependent pathway in adipocytes (in Chinese). *Acta Pharmacol Sin*, 1993, 31: 329–352 [方德秋, 章文才, 肖顺元. 应用同工酶进行柑桔分类和进化研究. 植物分类学报, 1993, 31: 329–352]
- 14 Zhong G Y, Ye M M. A numerical taxonomic study of *Citrus* and its close relatives. *Acta phytotaxonomica Sinica*, 1993, 31, 252–260 [钟广炎, 叶荫民. 柑桔植物的数值分类学研究. 植物分类学报, 1993, 31: 252–260]
- 15 Zhou Z Q, Jiang C Q. Study on the cladistic relationship of the genus *Citrus*. *South China Fruits*, 1994, 23: 3–6 [周志钦, 蒋聪强. 柑桔属植物若干分类群的分支分析. 中国柑桔, 1994, 23: 3–6]
- 16 Mabberley D. A classification for edible *Citrus* (Rutaceae). *Telopea*, 1997, 8: 167–172
- 17 Mabberley D J. A classification for edible citrus: an update, with a note on *Murraya* (Rutaceae). *Telopea*, 2022, 25
- 18 Zeng M. The understanding and opinion of the *Citrus* classification. *China Fruits*, 1962, 2: 31–37 [曾勉. 对柑桔分类的认识体会和整理意见. 中国果树, 1960, 2: 31–37]
- 19 Ollitrault P, Curk F, Krueger R. *Citrus* taxonomy, In Manuel T, Marco C, Fred C G, eds. The genus *Citrus*. Duxford: Woodhead publishing, 2020: 57–81
- 20 Samuel R, Ehrendorfer F, Chase M W, et al. Phylogenetic analyses of aurantioidae (Rutaceae) based on non-coding plastid DNA sequences and phytochemical features. *Plant Biol*, 2001, 3: 77–87
- 21 Hynniewta M, Malik S K, Satyawada R R. Karyological studies in ten species of *Citrus* (Linnaeus, 1753) (Rutaceae) of North-East India. *Comp Cytogenet*, 2011, 5: 277–287
- 22 Luo K, Chen S L, Chen K, et al. Assessment of candidate plant DNA barcodes using the Rutaceae family (in Chinese). *Sci Sin Viatae*, 2010, 4: 342–351 [[罗焜, 陈士林, 陈科. 等. 基于芸香科的植物通用 DNA 条形码研究. 中国科学: 生命科学, 2010, 4: 342–351]
- 23 Penjor T, Anai T, Nagano Y, et al. Phylogenetic relationships of *Citrus* and its relatives based on rbcL gene sequences. *Tree Genet Genomes*, 2010, 6: 931–939
- 24 Bao Y, Zeng Z, Yao W, et al. A gap-free and haplotype-resolved lemon genome provides insights into flavor synthesis and huanglongbing (HLB) tolerance. *Hort Res*, 2023, 10: uhad020
- 25 Huang Y, Xu Y, Jiang X, et al. Genome of a *Citrus* rootstock and global DNA demethylation caused by heterografting. *Hortic Res*, 2021, 8: 69

- 26 Wang L, He F, Huang Y, et al. Genome of wild Mandarin and domestication history of Mandarin. *Mol Plant*, 2018, 11: 1024–1037
- 27 Wang L, Huang Y, Liu Z A, et al. Somatic variations led to the selection of acidic and acidless orange cultivars. *Nat Plants*, 2021, 7: 954–965
- 28 Wang X, Xu Y, Zhang S, et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat Genet*, 2017, 49: 765–772
- 29 Wang X K, Jiang D, Sun Z Z. Study on phylogeny of 240 Mandarin accessions with genotyping-by-sequencing technology. *Sci Agr Sin*, 2017, 50: 1666–1673 [王小柯, 江东, 孙珍珠. 利用GBS技术研究240份宽皮柑橘的系统演化. *中国农业科学*, 2017, 50: 1666–1673]
- 30 Wu G A, Sugimoto C, Kinjo H, et al. Diversification of Mandarin *Citrus* by hybrid speciation and apomixis. *Nat Commun*, 2021, 12: 4377
- 31 Wu G A, Terol J, Ibanez V, et al. Genomics of the origin and evolution of *Citrus*. *Nature*, 2018, 554: 311–316
- 32 Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 2018, 34: i884–i890
- 33 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009, 25: 1754–1760
- 34 Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25: 2078–2079
- 35 McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 2010, 20: 1297–1303
- 36 Darriba D, Posada D, Kozlov A M, et al. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol*, 2020, 37: 291–294
- 37 Kozlov A M, Darriba D, Flouri T, et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 2019, 35: 4453–4455
- 38 Sanderson M J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 2003, 19: 301–302
- 39 Cejuela J M, Vinchurkar S, Goldberg T, et al. LocText: relation extraction of protein localizations to assist database curation. *BMC BioInf*, 2018, 19: 15–30
- 40 Pickrell J, Pritchard J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nat Prec*, 2012: 1–1
- 41 Durand E Y, Patterson N, Reich D, et al. Testing for ancient admixture between closely related populations. *Mol Biol Evol*, 2011, 28: 2239–2252
- 42 Huang Y, He J, Xu Y, et al. Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in *Citrus* fruits. *Nat Genet*, 2023, 55: 1964–1975
- 43 Hodgson R W. Taxonomy and nomenclature in *Citrus*. International Organization of Citrus Virologists Conference Proceedings, Gainesville. 1961. 1–7
- 44 Pang X M. Studies on phylogeny of *Citrus* and its relatives and on genetic diversity of Poncirus using molecular markers (In Chinese). Dissertation for Doctoral Degree. Wuhan: Huazhong Agricultural university, 2002 [庞晓明. 用分子标记研究柑橘属及其近缘属植物的亲缘关系和种的遗传多样性. 博士学位论文. 武汉: 华中农业大学, 2002]
- 45 Wu Y U, Liu Q, Xiang Z H, et al. Citrus × pubinervia, a new natural hybrid species from central China. *Phytotaxa*, 2021, 523: 239–246
- 46 Carbonell-Caballero J, Alonso R, Ibañez V, et al. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol Biol Evol*, 2015, 32: 2015–2035
- 47 Yang X, Li H, Yu H, et al. Molecular phylogeography and population evolution analysis of *Citrus ichangensis* (Rutaceae). *Tree Genet Genomes*, 2017, 13: 29
- 48 Garcia-Lor A, Luro F, Ollitrault P, et al. Genetic diversity and population structure analysis of mandarin germplasm by nuclear, chloroplastic and mitochondrial markers. *Tree Genet Genomes*, 2015, 11: 1–5
- 49 Li Y Z. Research on the genetic diversity and phylogenetic relationship of loose skin mandarins (*Citrus reticulata* Blanco). Dissertation for Doctoral Degree. Wuhan: Huazhong Agricultural university, 2006 [李映志. 宽皮柑橘(*Citrus reticulata* Blanco)的遗传多样性及系统发育研究. 博士学位论文. 武汉: 华中农业大学, 2006]
- 50 Xiao H, Liu Z, Wang N, et al. Adaptive and maladaptive introgression in grapevine domestication. *Proc Natl Acad Sci USA*, 2023, 120: e2222041120
- 51 Sun X, Jiao C, Schwaninger H, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*, 2020, 52: 1423–1432
- 52 Guo S, Zhao S, Sun H, et al. Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat Genet*, 2019, 51: 1616–1623

The definition, classification and evolution of citrus based on whole genome sequence information

XU Qiang, HUANG Yue & DENG XiuXin

National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, Huazhong Agricultural University, Wuhan 430070, China

Citrus is one of the most widely grown economic crops in the world. The wide sexual compatibility on inter-specific and even inter-generic level and the prevalent asexual variation of *Citrus* have led to the formation of the highly diverse and complex genetic background of citrus species. The classification of *Citrus* has also become one of the most challenging topics in plant science. In this study, we collected the most comprehensive genome sequencing data of *Citrus* so far for a precise analysis of the genetic composition and proportions of various varieties. We divided mandarin, pummelo and their hybrids into five types, and assigned a definition for each group. Then, using “Orah” and “Beni Madonna” as examples, we analyzed the genomic changes along with the breeding history of *Citrus*, and proposed the dispersal route of mandarin after domestication. We discussed the classification and origin of *Citrus* species and the genetic differences between “gan” and “ju”, and also discussed the migration and evolution of mandarin. We proposed an objective classification of *Citrus* based on whole genome information, which provided a framework for the utilization of germplasm resources.

***Citrus*, classification, origin, genome**

doi: [10.1360/SSV-2023-0229](https://doi.org/10.1360/SSV-2023-0229)